Mediation Centrality in Adversarial Policy Networks

Stefan M. Herzog
Center for Adaptive Rationality
Max Planck Institute for Human Development, Berlin
(herzog@mpib-berlin.mpg.de)

Thomas T. Hills
Department of Psychology
University of Warwick
(t.t.hills@warwick.ac.uk)

Conflict resolution often involves mediators who understand the issues central to both sides of an argument. Mediators in complex networks represent nodes that are connected to other key nodes in opposing subgraphs. Here we introduce a new metric, *mediation centrality*, for identifying good mediators in adversarial policy networks, such as the connections between individuals and their reasons for and against the support of controversial topics (e.g., state-financed abortion). Using a process-based account of reason mediation we construct bipartite adversarial policy networks and show how mediation defined over subgraph projections constrained to reasons representing opposing sides can be used to produce a measure of mediation centrality that is superior to centrality computed on the full network. We then empirically illustrate and test mediation centrality in a "policy fluency task," where participants generated reasons for or against eight controversial policy issues (state-subsidized abortion, bank bailouts, forced CO_2 reduction, cannabis legalization, shortened naturalization, surrogate motherhood legalization, public smoking ban, and euthanasia legalization). We discuss how mediation centrality can be extended to adversarial policy networks with more than two positions and to other centrality measures.

Introduction

Adversarial systems can be defined as systems composed of individuals with opposing views, such as Democrats versus Republicans in US politics or Leave versus Remainers in the Brexit discussion. Numerous recent studies have investigated the development of adversarial information environments that can isolate individuals from the views of their opponents, such as echo chambers and filter bubbles (Barberá et al., 2015; Nikolov et al., 2015; Weng et al., 2012). This isolation can lead to overconfidence and further polarization and, counter-intuitively, may be especially prominent in information-rich environments (Hills, 2018). These systems often form over ideological divisions and extend even to the truth value of science, with the end result that such groups rarely see eye-to-eye and are severely insulated from one another in relation to beliefs and social contacts (e.g., Fiorina & Abrams, 2008; Shi et al., 2017).

To make progress on controversial issues in adversarial systems, it can be useful to identify individuals who are best able to help collective problem solving. Such individuals should be able to guide others towards recognizing and acknowledging the beliefs and values of individuals on different sides of an issue (Mutz, 2002). For example, a capacity for perspective taking—the ability to understand and acknowledge views on different sides of an issue—is one of the most effective tools of a good negotiator (Galinsky et al., 2008). Similarly, convergent framing that identifies a collectively recognized description of the problem can help facilitate conflict resolution (Drake & Donohue, 1996). From the perspective of conflict resolution (Deutsch et al., 2006), it is this ability to recognize the collective perspectives (Ku et al., 2015) that is a defining characteristic of a good *mediator*. Such good mediators maximize the opportunity that the majority of individuals on each side of the issue can agree on what the disagreement really comes down to.

Adversarial systems of the kind described above can be considered complex networks, where individuals are connected to other individuals by acknowledgement of shared reasons supporting opposing sides of an issue. Although network science has proposed many centrality metrics for identifying key nodes in a variety of contexts (Bonacich, 2007; Borgatti, 2005; Freeman, 1977; Newman, 2018; Sabidussi, 1966), we know of no metric for identifying mediators in adversarial systems, and more specifically adversarial policy networks, where a network is adversarial because it contains information representing more than one position about the policy. Adversarial policy networks can be represented by bipartite networks, where individuals are connected by edges to the reasons they acknowledge. Figure 1A provides an empirical example of such a bipartite adversarial policy network based on reasons—and the individuals who recognize those reasons—concerning the policy issue of reducing the minimum number of years of residence to become a naturalized citizen in Switzerland (based on empirical results of a study described later in the paper). In such adversarial policy networks, the reasons can support only one of several sides of an issue. Such a network can be projected onto persons (where

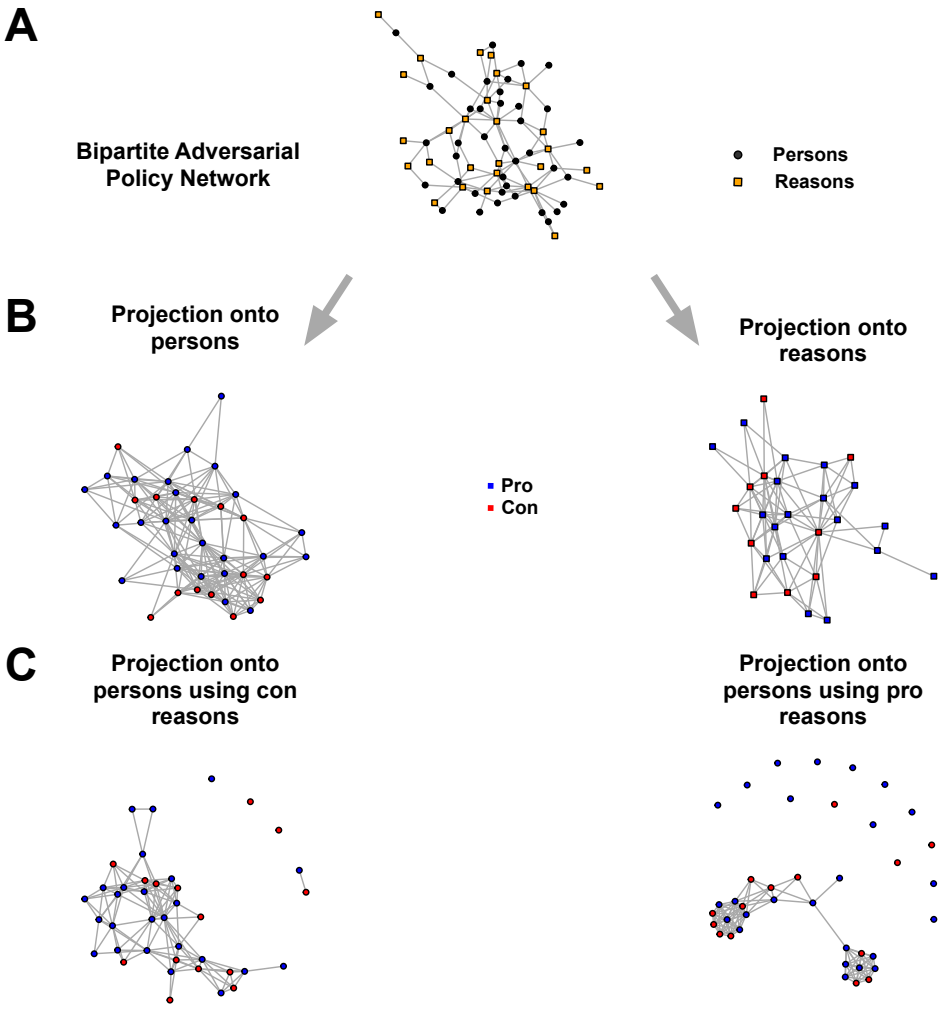

Figure 1. An example of a bipartite adversarial policy network around the issue of speeding up naturalization of foreigners in Switzerland (based on empirical results of a study described later in the paper). **(A)** The bipartite network composed of persons and the reasons they acknowledge. **(B)** Projections of the bipartite network onto persons, where individuals are connected if they share at least one reason, or onto reasons, where reasons are connected if they are both produced by the same person. Blue indicates reasons or persons favoring reduced naturalization times and red indicates reasons or persons against reduced naturalization times. **(C)** The person projections based on bipartite networks constrained to only contain either pro or con reasons. Blue indicates persons favoring reduced naturalization times and red indicates persons against reduced naturalization times; persons' attitudes towards the policy issue were assessed using a separate survey item.

individuals are connected if they share at least one reason) or reasons (where two reasons are connected if they are produced by the same person; Figure 1B). By constraining the reasons to be on one side of the issue one can further describe subgraphs of individuals who are connected in relation to either pro or con reasons (Figure 1C).

In bipartite adversarial policy networks, there are reasons for and against the policy as well as individuals who acknowledge different subsets of those reasons, with some individuals recognizing reasons on both sides of the issue. A good mediator in this space is someone with high centrality in all subgraphs. Importantly, by this definition, a good mediator is not necessarily someone who recognizes the most reasons or the reasons that would make the most people happy, nor is it someone who acknowledges an equal amount of reasons among the various positions—or even a person who rec-

ognizes those reasons that would best cover the reason space as defined by what people collectively acknowledge (a metric, representativeness, which we describe below). Each of these attributes can be gamed simply by adding more people or poor quality reasons. The method we describe below is immune to such subterfuge.

The central contribution of this paper is the introduction and investigation of *mediation centrality*—a network measure for identifying mediators in bipartite adversarial policy networks. Mediation centrality is computed by combining centrality metrics from subgraph projections where the projections are defined in relation to different sets of reasons.

In what follows, we first define bipartite adversarial policy networks and then describe our novel mediation metric for identifying graph mediators on these networks. We then evaluate this mediation metric using simulated bipartite adversarial policy networks and show that mediation centrality captures the notion of a good mediator who is the best-recognizer-of-best-recognized-reasons in a hypothetical discussion that follows an associative path through the argument space. A good mediator in this space would have the most to contribute to this discussion. We then empirically illustrate this mediation metric in a "policy fluency task," where participants generated reasons for or against a range of controversial policy issues.

Mediation centrality

Bipartite adversarial policy networks, mediation, and centrality metrics

Bipartite adversarial policy networks can be represented as graphs, $G(V, E)$, where vertices, V, are composed of individuals, I, and reasons, R, with edges, E, connecting individuals to reasons (Figure 1A). For the network to be adversarial, reasons represent positions with respect to the policy and are therefore exclusive to one subgraph. As we note below, this can be extended to any number of positions, but for ease of exposition we assume that reasons can only be either *for* or *against* the policy.

The projection of reasons onto individuals gives a graph $G_I(I, E)$ where individuals i and j are connected in the resulting adjacency matrix if they share at least one reason, $k \in R$, (Figure 1B, left graph), such that

$$A_{ij} = \sum_k R_{i,k} R_{j,k}$$

where $R_{i,k}$ has a value of 1 if reason k is held by individual i and 0 otherwise. Similarly, one can form projections onto reasons (Figure 1B, right graph).

The different sides of the position can be represented by $G_+(I, R_+; E)$ and $G-(I, R_-; E)$, representing the subgraphs formed by constraining reasons to those either for or against the issue, respectively. Forming the projections onto individuals as above, we get $G_{I,+}$ and $G_{I,-}$, respectively, which represent individuals' connectivity solely driven by either pro or con reasons, respectively (Figure 1C).

There are a variety of centrality metrics that could be computed on each of the subgraph projections, such as degree centrality, betweenness centrality, and closeness centrality. Mediation centrality can be generalized to each of these metrics as we will discuss below. However, because we are considering mediation in the context of a domain where reasons are represented in individuals' minds, we are interested in how ideas are connected between people. In particular, we are interested in the process of a hypothetical fruitful discussion, where the discussion tracks the structural information defined by associations between people and the reasons they collectively acknowledge. In such a setting, a good mediator is someone who would contribute maximally to this hypothetical discussion because she knows and can introduce the collectively important reasons into the discussion. She therefore is a best-recognizer-of-best-recognized reasons (as we show below). We identify this mediator by making two assumptions: that thought is associative and that people are connected, among other things, by shared ideas.

In "Trayne of Thoughts," (Hobbes, 1998) recognized what has become a truism in contemporary cognitive science: one thought gives rise to another in relation to the association between them (Hills et al., 2012; Griffiths et al., 2007). Extending this idea to a collection of individuals in an adversarial policy network, we imagine a simple model of a social process whereby individuals activate one another by their shared reasons, with one individual stating one reason and another responding to that reason with another associated reason that comes to mind. This process can be formally described as a random walk through policy space, where transitions between individuals occur by choosing edges at random in the collective associative representation.

A projection of the bipartite adversarial policy network onto individuals captures this process, whereby individuals are connected if they can activate one another through a shared idea. A random walk over this subspace is equivalent to the process described above. The probability of moving between two nodes in this subspace is described by the transition matrix $\mathbf{T}$ and describes a Markov process which converges to a stationary distribution over successive transitions (Norris, 1998).

This stationary distribution can be represented by the vector $\mathbf{x}$. Stationarity implies that further transitions do not affect the distribution, such that

$$\mathbf{x} = \mathbf{T}\mathbf{x},$$

where $\mathbf{x}$ is the eigenvector associated with the largest eigenvalue of $\mathbf{T}$. $\mathbf{T}$ is the normalized adjacency matrix represented

as follows,

$$\mathbf{T_{ij}} = \mathbf{A_{ij}} / \sum_{\mathbf{k=1}}^{\mathbf{n}} \mathbf{A_{kj}}$$

The values of the stationary distribution $\mathbf{x}$ corresponds to a special case of *PageRank* (Brin & Page, 1998) for nondirected graphs. PageRank is a network measure, which has been applied to numerous cognitive and social phenomena (e.g., Austerweil et al., 2012; Borge-Holthoefer & Arenas, 2010; Ding et al., 2009; Griffiths et al., 2007). Roughly speaking, the PageRank of a node corresponds to the probability of finding a random walker at that node, where the walker is subject to a Markov process constrained by the adjacency matrix. Although mediation centrality can be generalized, in principle, to any centrality metric (such as degree centrality, betweenness centrality, and closeness centrality), in the description of mediation centrality that follows we will restrict our investigation to PageRank because it follows the logic outlined above of a discussion constrained by the structure of associative relations between reasons as they occur among people. However, in the discussion we will argue that which centrality metric is most appropriate for any domain will depend on the processes involved in that domain.

Computing mediation centrality over subgraphs

Centrality measures are routinely computed on the full network and may therefore be considered global centrality measures. As we show later, a global measure of centrality does not capture the basic logic of a good mediator because mediators need to be able to mediate discussions between different positions. That is, in the context of adversarial policy networks, an individual with high centrality on $G_I(I, E)$ may not be a good mediator across opposing subgraphs. Specifically, they may not acknowledge issues that would make them central to G_{I-} and G_{I+} at the same time. To handle this problem, we define a node's mediation centrality, M, as the harmonic mean of its centrality values across subgraphs

$$M_i = \frac{2}{1/x_{i,+} + 1/x_{i,-}} = \frac{2x_{i,+}x_{i,-}}{x_{i,+} + x_{i,-}}$$

where $x_{i,+}$ and $x_{i,-}$ represent the centrality computed for node i from the subgraphs G_{I+} and G_{I-}, respectively. The harmonic mean captures our intended notion of mediation centrality because it is dominated by the smallest (minimum) centrality across the subspaces. In particular, the right-most part of the above equation for M_i highlights that if either $x_{i,+}$ or $x_{i,-}$ is zero, $M_i = 0$—irrespective of the value for the other centrality.

More generally, the harmonic mean H is a Schur-concave function, which implies that for any positive set of inputs we have $\min(x_1 \dots x_n) \leq H(x_1 \dots x_n) \leq n \min(x_1 \dots x_n)$. This means that H cannot be made arbitrarily large without also changing the value of its smallest input. In particular, if any input is zero, $H = 0$, irrespective of the values of all other inputs. The geometric mean is also a Schur-concave function. However, we chose the harmonic mean in analogy to computations of average speed: When a vehicle travels at rate a and then at rate b for equal distances, then the average rate is the harmonic mean of a and b. Loosely speaking, the average "rate" of a mediator's contribution within a subspace is how often they contribute to the random walk in the associated subgraph (i.e., how often an individual is visited by the random walk). If we assume an analogous concept of equidistant paths through the pro and con argument spaces, the average rate of an individual's contribution is proportional to the harmonic mean of their contributions over subgraphs.

For adversarial policy networks consisting of n subspaces representing n positions, mediation centrality is defined as

$$\mathbf{M} = \frac{\mathbf{n}}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i}^{-1}},$$

where $\mathbf{x_i}$ is the centrality for the ith subspace computed from $G_{I,i}$. This conveniently reduces to the standard centrality measure for the case of only one subspace ($n = 1$).

Note that mediation centrality is different from the subgraph centrality described by Estrada & Rodriguez-Velazquez (2005), which merely "counts the times that a node takes part in the different connected subgraphs of the network," such as triangles, four cycles, and so on.

Mediation centrality vs. representativeness

Earlier we argued for a process-based measure of mediation that can capture structural relations between social and cognitive processes. We then identified PageRank as a suitable network metric for such a mediation measure. Here we introduce and formalize the cognitive notion of *representativeness*, a cognitive measure designed to capture reason coverage in a population. We note that representativeness—unlike mediation centrality—does not incorporate network structure. Comparing mediation centrality to representativeness helps to highlight the potential weaknesses of simple counting measures.

To develop the notion of representativeness, let us assume a unit called a *person–reason*, which represents a reason held by one person. Two person–reasons can reflect two different people who each acknowledge one reason—which may or may not be the same reason—or one individual who acknowledges two different reasons. By this unit, ten people who all acknowledge the same, one reason (= 10 person–reasons) have a less formidable reason space than ten people who acknowledge five reasons each (whether or not they are shared; = 50 person–reasons). If we assume that a reason in an individual's mind represents a slot in the adversarial space, then the total adversarial space for one side of an issue is the sum of the slots, that is, the sum of all person–reasons. This assumes that reason slots are interchangeable. This may

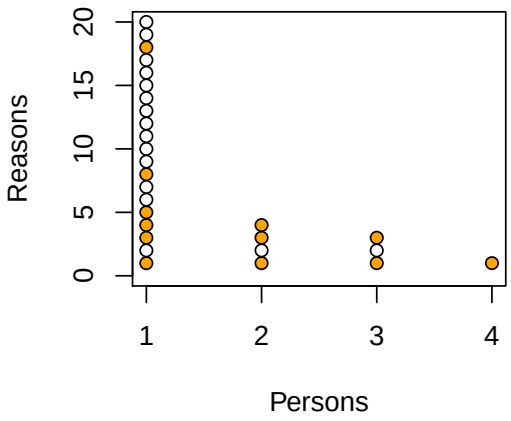

Figure 2. The reason space consists of reasons and the number of individuals that hold those reasons. We present a random individual who acknowledges reasons #1, #3, #4, #5, #8, and #18, which are held by 4, 3, 2, 1, 1, and 1 persons, respectively. The person-reasons covered by this person are shown in orange and represent this individual's representativeness, $\rho = 4 + 3 + 2 + 1 + 1 + 1 = 12$.

not always be the case as some reasons may be more convincing than others even though they are held by fewer individuals. Incorporating a reason's normative weights is beyond the scope of the current paper (but see Russell & Reimer, 2018, for some ideas to build on).

For illustration, Figure 2 depicts an adversarial space consisting of 20 unique reasons and 28 person–reasons—with one reason held by 4 individuals, two reasons held by three individuals each, and so on. The extent to which an individual covers this space is a measure of their representativeness ρ. We can therefore define the representativeness of an individual as the sum of the slots (person–reasons) they cover in the adversarial space

$$\rho = \sum_{i \in C}^{n} w_i,$$

where for each reason i in that individual's set of reasons, C, we sum the number of individuals w_i who acknowledge that reason. An individual with a higher ρ is an individual who better covers the adversarial space over which C is represented. Accordingly, we can compute ρ_p and ρ_c to indicate the representativeness within the pro and con reason spaces, respectively.

As computed here, a node's representativeness within a subgraph is equivalent to the node's weighted degree in a network, where edge weights reflect the number of shared reasons, plus the nodes unweighted degree, representing the number of reasons held by the individual. The addition of the node's unweighted degree could be removed to avoid situations where an individual creates unique, idiosyncratic reasons to amplify their own representativeness. In the present case, we leave this in with the assumption that the individuals in the network reflect the population from which they are sampled. This also fits the framing in the empirical study below where we asked individuals to generate reasons they believe would be held by others.

We define *mediation representativeness* over multiple spaces, $\hat{\rho}$, by the harmonic mean of the representativeness over subgraphs,

$$\hat{\rho} = \frac{n}{\sum_{i=1}^{n} \rho_i^{-1}}$$

where n is the number of subspaces. As in the definition of median centrality (see earlier), the harmonic mean best captures our intended meaning of mediation representativeness because it is dominated by the smallest representativeness across the subspaces. This prevents individuals from becoming more representative by merely capturing a larger share of an already well-represented subspace.

Though M and $\hat{\rho}$ may often be correlated in practice—and indeed are well correlated in the empirical study we describe below—they need not be correlated. To see why, consider a reason network where two candidate mediators have identical Ms and $\hat{\rho}$s. The first candidate can improve her $\hat{\rho}$ by listing one more reason on each side of the policy issue. However, her M remains unchanged, as it crucially depends on the reasons being recognized by others. In other words, M does not change because it is sensitive to the structure of the person–reason network space.

Evaluation: Simulation studies

Simulation 1

Consider a policy debate with two positions, for and against, with corresponding *pro* and *con* reasons. Individuals are aware of various reasons on both sides of the debate. The goal is to identify mediators in this space who are the best recognizers of best-recognized reasons on both sides of the debate.

To simulate this, let there be $N = 100$ individuals in a policy debate on an issue (e.g., legalization of cannabis) where there is a universe of 10 possible distinct *pro* reasons and 10 possible distinct *con* reasons. Each individual samples a total of 10 reasons: 10β *pro* reasons and $10\,(1 - \beta)$ *con* reasons (rounded to the nearest integer). β then represents an individual's *bias*; a $\beta = 0.5$ represents an unbiased individual. In this simulation, an individual's β is uniformly sampled from $[0, 1]$.

We allow reasons to have a power law distributed probability p of being sampled, where $p \sim r^{-\gamma}$ with $\gamma = 2.5$ and

rank r. The precise value of γ is unimportant to the overall results except that for larger values of γ all individuals will produce the same reason and for small values of γ all reasons will be sampled uniformly. In such cases everyone is the best mediator (since everyone produces the same reasons) or the best mediator is the individual with the most number of reasons per position, since all reasons are equally represented. Thus, intermediate γ's are the most interesting formulation and also the formulation that best reflects the empirical data presented later.

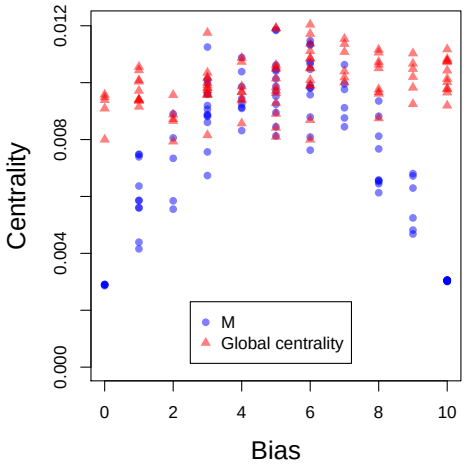

Figure 3. Individuals' centrality values and bias in Simulation 1. The centrality value of an individual is computed using two methods (mediation centrality and global centrality) and shown on the y-axis as a function of the individual's bias. Bias is shown here as the integer number of reasons on the pro side (i.e., 10β).

Using what individuals sample, we produce a bipartite adjacency matrix where individuals are rows and reasons are columns. As described above, we then project this matrix onto individuals for the pro and con reason subspaces separately and then compute individuals' mediation centrality as the harmonic mean of their centrality measures across both subspaces.

Figure 3 shows the relationship between global centrality (i.e., computed from the projection onto individuals based on the full bipartite network) and mediation centrality. Global centrality shows a limited ability to discriminate among individuals' different degrees of bias for or against the policy issue. Mediation centrality, on the other hand, captures the central intuition of mediation, whereby individuals with the least bias, $\beta = .5$, have the highest mediation centrality. Notably, however, the individual with the smallest bias is not necessarily the one with the highest mediation centrality. The variation in mediation for a given level of bias is a measure of the individual's ability to capture the most well-recognized

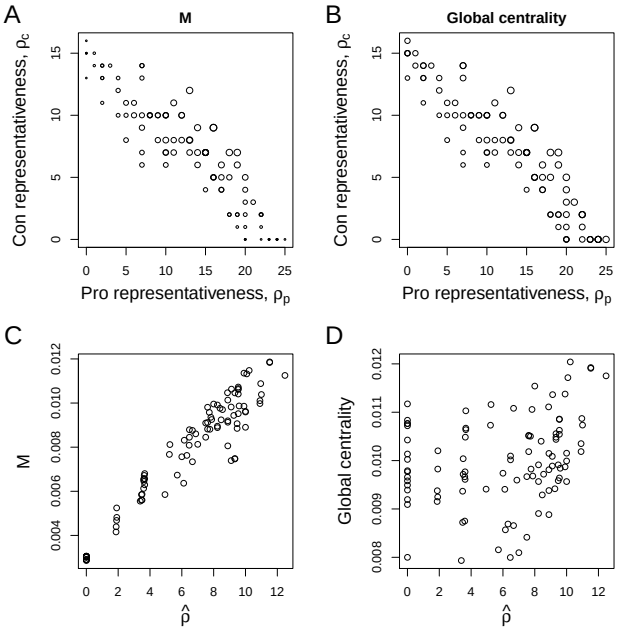

Figure 4. Comparison of mediation representativeness against mediation centrality and global centrality in Simulation 1. Each individual is represented by a dot. **(A)** Representativeness for pro and con reasons. Dot size represents mediation centrality value. **(B)** Representativeness for pro and con reasons. Dot size represents global centrality. **(C)** The global representativeness against mediation centrality. **(D)** The global representativeness against global centrality.

reasons within a subgraph. Consider that individual's with no bias ($\beta = 0.5$) will produce an equal number of reasons from both sides of the issue, but these reasons may not be equally representative of the sides from which they are sampled. Therefore, more biased individuals can (up to a point) still be better recognizers of best-recognized reasons.

Figure 4 relates each simulated agent's representativeness ρ for pro and con spaces against each other and against the same agent's mediation centrality, M, and global centrality. The figure reveals a number of insights. Foremost, as to be expected, individuals who have higher representativeness in the con subspace have lower representativeness in the pro subspace. This is because the number of reasons produced is fixed and split across the subspaces. Secondly, individuals who are highly representative of one or the other of the two subspaces have lower mediation centrality (smaller sized dots; Figure 4A). Global centrality, on the other hand, is influenced by greater representativeness of either pro or con sides (Figure 4B). A comparison of mediation centrality with representativeness across both the pro and con spaces, $\hat{\rho}$, shows that mediation centrality captures the intuition that mediators must acknowledge key ideas on each side of the issue (Figure 4C); global centrality lacks this property (Fig-

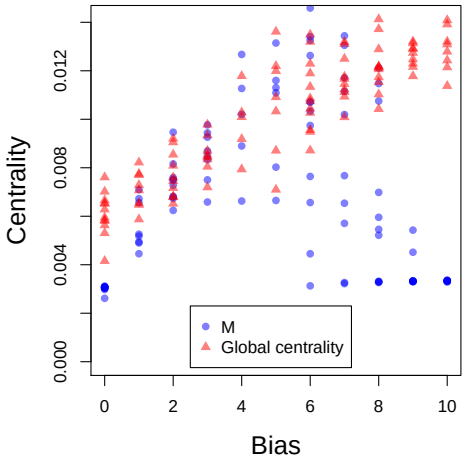

Figure 5. Individuals' centrality and bias values in Simulation 2. The centrality value of an individual is computed using two methods (mediation centrality and global centrality) and shown on the y-axis as a function of the individual's bias.

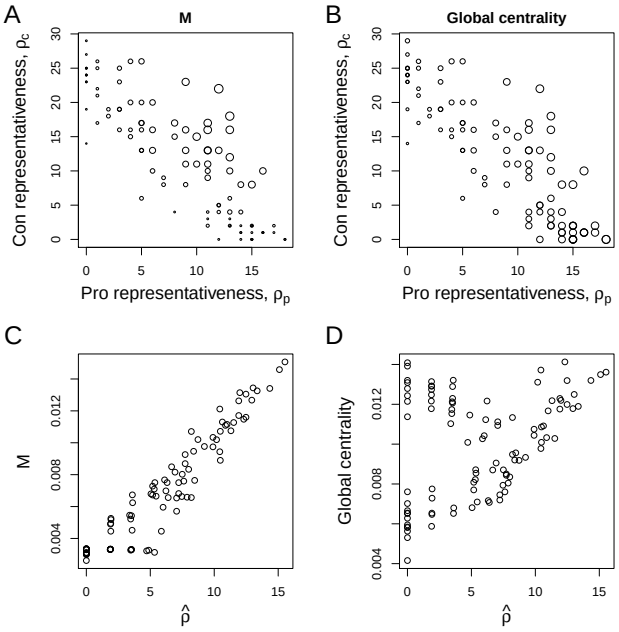

Figure 6. Comparison of mediation representativeness against mediation centrality and global centrality in Simulation 2. Otherwise this figure follows Figure 4.

ure 4D). In addition, also note that the individual with the highest M is not the individual with the highest $\hat{\rho}$.

Following the notion of a random walk through policy space, we can also verify that mediation centrality tracks the residence time of random walkers on each subgraph. Figure A1 in the Appendix shows the outcome of releasing 1,000 random walkers from each individual and tracking the residence times for each node in the network. Mediation centrality is again highest for the individuals who are least biased in their residence times (Figure A1-A). Global centrality, in contrast, is less discriminating (Figure A1-B). Comparing mediation centrality against the harmonic mean of residence times shows a close relationship between the two measures (Figure A1-C). Global centrality, in contrast, does not show a clear relationship with the harmonic mean of residence times (Figure A1-D).

Simulation 2

To further examine the characteristics of a good mediator, we ran a second simulation where there are 100 reasons in total, 80 against and 20 in favor of the policy issue; all other simulation details are identical to Simulation 1. Figures 5, 6, and A2 (in the Appendix) show the corresponding results. As expected, mediation centrality is relatively unaffected by the imbalance between the number of con and pro reasons and identifies individuals who are both unbiased and more representative; global centrality, in contrast, cannot capture the differences between the two subgraphs (Figure 5). Note also that the individual with the highest global centrality has the lowest mediation representativeness and is strongly biased towards producing pro reasons (Figure 6D). Finally, mediation centrality again tracks the harmonic mean of random walker residence times, whereas global centrality fails to capture this (Figure A2 in the Appendix).

The two simulations presented have highlighted the usefulness of mediation centrality in identifying good mediators across two sides of a policy issue. In the next section, we apply this measure to empirical data on actual policy issues.

Mediators in Adversarial Policy Networks: An Empirical Study

To test mediation centrality in a real-world context, we collected data for eight policy issues (Table 1). Participants in this study were asked to imagine that they would be moderating a discussion on a specific policy proposition and that in preparation for this meeting they should list all the possible reasons in favor or against the proposition they could think of that might come up in such a discussion. We call this task the "policy fluency task" following similar tasks in the category fluency literature, such as the animal and country fluency tasks, where individuals name all the animals or countries they can think of, respectively (e.g., Hills et al., 2012; Hills & Segev, 2014).

Methods

Participants. Fifty-three participants (median age = 22; 40 females) were recruited at the University of Basel

Policy issue	Policy question: Should...	Con	Pro	% Pro
State-subsidized abortion	... the state subsidize abortions?	33	18	35
Bank bailouts	... the state bail out banks during an economic crisis?	20	26	57
Forced CO_2 reduction	... developing countries be forced to reduce CO_2 emissions?	29	20	41
Cannabis legalization	... the possession and consumption of cannabis be legalized?	29	20	41
Shortened naturalization	... the minimum years of residency for citizenship be reduced?	14	25	64
Surrogate legalization	... surrogate motherhood be legalized?	25	23	48
Public smoking ban	... public smoking be banned?	23	28	55
Euthanasia legalization	... medically assisted suicide by legalized?	22	31	58

Table 1

The 8 Policy issues. Con and Pro: Number of participants in favor of or against the policy; % Pro: percentage of participants in favor of the policy. The issues were framed in the context of the country in which the study was conducted (Switzerland).

(Switzerland). As this experiment was a non-clinical study and did not involve any patients, according to Swiss federal law it did not require an in-depth evaluation and approval by a cantonal review board.

Materials and Procedure. We conducted a pilot survey to identify policy issues for which in our participant population there is non-negligible support for both sides of an issue. Table 1 shows the eight issues used in the main study.

The primary data for this study are the reasons participants generated for each policy issue. Next to this primary data, we collected several additional variables that were not investigated in relation to mediation centrality. In the spirit of full disclosure we nevertheless report them below when describing the experiment. All instructions were in German; we present their English translations here. The experiment was programmed in E-Prime 2.

1. To measure working memory capacity, participants completed an operation span task (Unsworth et al., 2005).

2. Participants were asked to imagine that they would be mediating a discussion on a specific policy proposition (e.g., legalizing cannabis) and that their role was that of an impartial mediator. In preparation for this discussion they would list all the arguments (i.e., reasons) for and against the current proposition they could think of that other people might find important for deciding in favor or against the policy proposition. For each of the eight policy issues (Table 1), participants were instructed to write down each reason they could think of using 3-4 words and submit it by pressing ENTER. Once they could not think of any more reasons, they proceeded to the next issue by pressing a button on the screen. Policy issues were presented in a new random order for each participant.

3. For each of the eight policy issues participants indicated their own position (i.e. in favor or against the policy proposition).

4. Participants indicated a set of demographic variables (age; gender; Swiss citizen status; smoking status; cannabis consumption).

5. Several self-rating questions assessed participants' political stance. The first question asked participants to place themselves on the political left- vs. right-wing spectrum by choosing a point on an analogue scale. Then for each of eight Swiss political parties participants indicated to which degree they agreed or disagreed with their political agenda. Participants choose a point on an analogue bipolar scale that ranged from "total disagreement" to "total agreement". The eight political parties were: Schweizer Volkspartei (SVP); Sozialdemokratishe Partei (SP); Freisinning demokratische Partei/FDP - Die Liberalen; Christlichdemokratische Volkspartei (CVP); Grüne Partei (GPS); Bürgerlich-Demokratische Partei (BDP); Grünliberale Partei (GLP); Evangelische Volkspartei (EVP)

6. We assessed participants' self-reported ability for perspective taking based on the four items of the German version (Paulus, 2009) of the Interpersonal Reactivity Index (Davis, 1983). Participants indicated the degree to which four statements applied to them by choosing a point on an analogue bipolar scale that ranged from "does not apply at all" to "fully applies". In their original English formulation (Davis, 1983) the statements read: "I try to look at everybody's side of a disagreement before I make a decision." "I believe that there are two sides to every question and try to look at them both." "Before criticizing somebody, I try to imagine how I would feel if I were in their place." "When I'm upset at someone, I usually try to 'put myself in his shoes' for a while."

Three raters independently judged for each reason whether it was in support of (+1) or against (−1) the policy issue or whether they could not tell (0); we then summed the values and took the valence of the sum to indicate whether

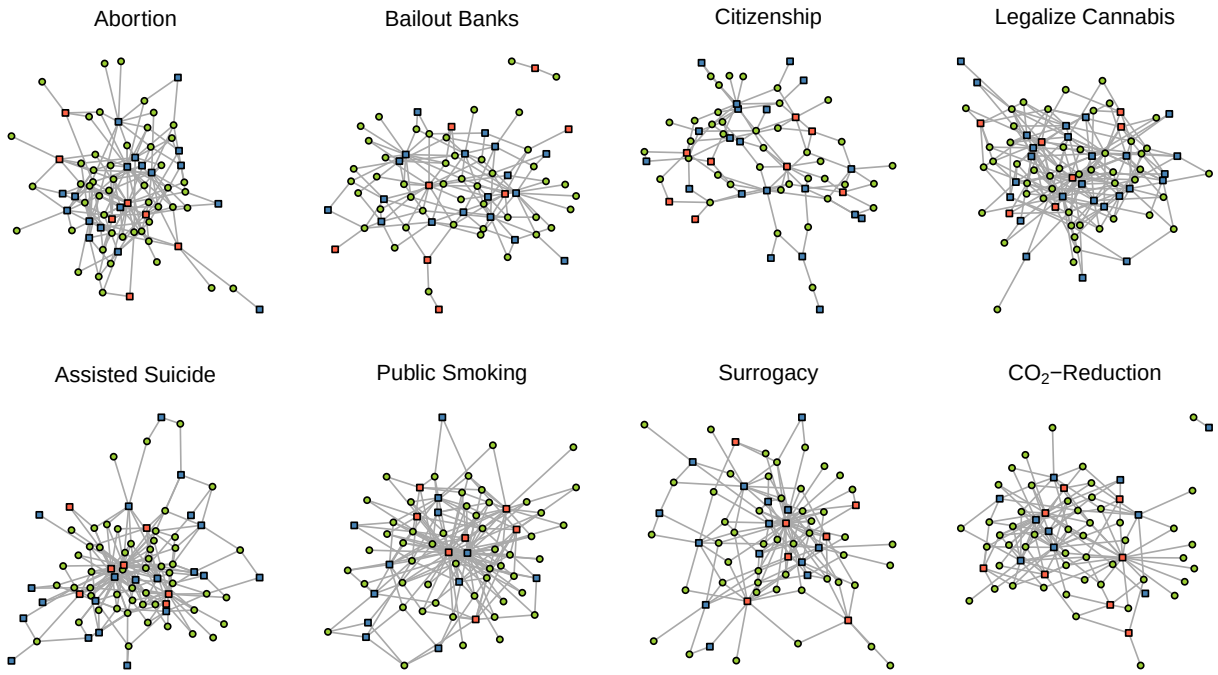

Figure 7. Bipartite adversarial policy networks for each issue, with reasons shown as boxes (red = con, blue = pro) and individuals shown as circles (green).

the reason was pro or con. Out of the total 1,778 reasons produced, 324 had zero valence. We excluded these as they often do not refer to coherent reasons.

For each issue, a fourth rater created overarching categories of reasons to which the produced reasons were then assigned; 7 of the remaining 1,454 reasons failed to be coded and were removed. The assigned categories were then used to compute the values in the adjacency matrices. For example, one individual wrote "murder of the fetus" and another wrote "it is murder" which were then classified under the same category "abortion is murder." An individual's representativeness was calculated by considering the number of unique reason categories for which a participant produced at least one reason. This was done to avoid inflated values of representativeness when a participant produced multiple reasons that all belonged to the same reason category.

After the two rating procedures, 1,447 reasons remained, which were used in all further analyses reported. The sum of the counts in Table 1 indicate how many out of the 53 participants produced at least one valid reason for the respective issue.

Results

Table 1 shows the number of individuals on each side of each issue. Our pilot survey aimed to identify policy issues for which there would be substantial support for either side in our participant population. Consistent with this goal, each of the issues showed non-negligible support for both positions. These levels of polarization suggest that the issues used in the study represent a good test bed for investigating mediation centrality.

Figure 7 shows the bipartite adversarial reason networks for each issue. The networks each have one giant component, which shows that people on both sides of each issue tend to acknowledge reasons on both sides of the issue. Thus, even though these are controversial issues, participants were—at least partly—aware of the reasons the other side holds. This implies that identifying mediators as individuals who are the best-recognizers-of-best-recognized-reasons is a plausible endeavor in this study.

Mediation centrality is useful to the extent that it varies across individuals in adversarial policy networks. Figure 8 shows that mediation centrality produces a clear ranking of individuals within each of the eight very different policy issues. This is promising for two reasons. First, it implies that even among controversial topics there is a range of adversarial understanding among people. Second, more pragmatically, the result also implies that there are individuals who would be much better mediators than others.

Although the mere number of reasons produced by an participant is a rough proxy for the participant's mediation centrality, it is a poor direct substitute for mediation centrality. Figure 9 shows that although the highest mediation centrality corresponds in some cases to the individual with

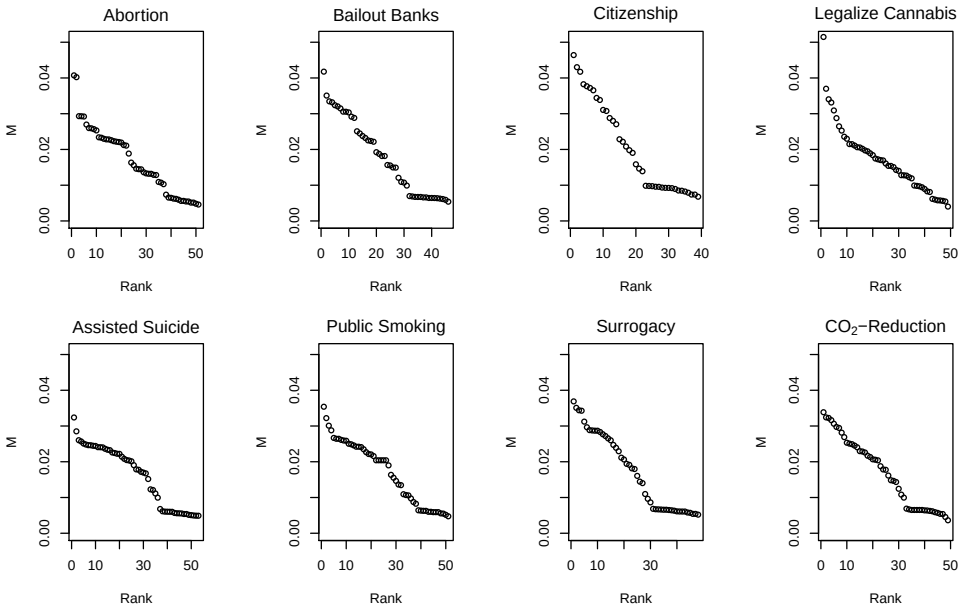

Figure 8. Participants' mediation centrality, ranked from largest to smallest, for each of the eight policy issues.

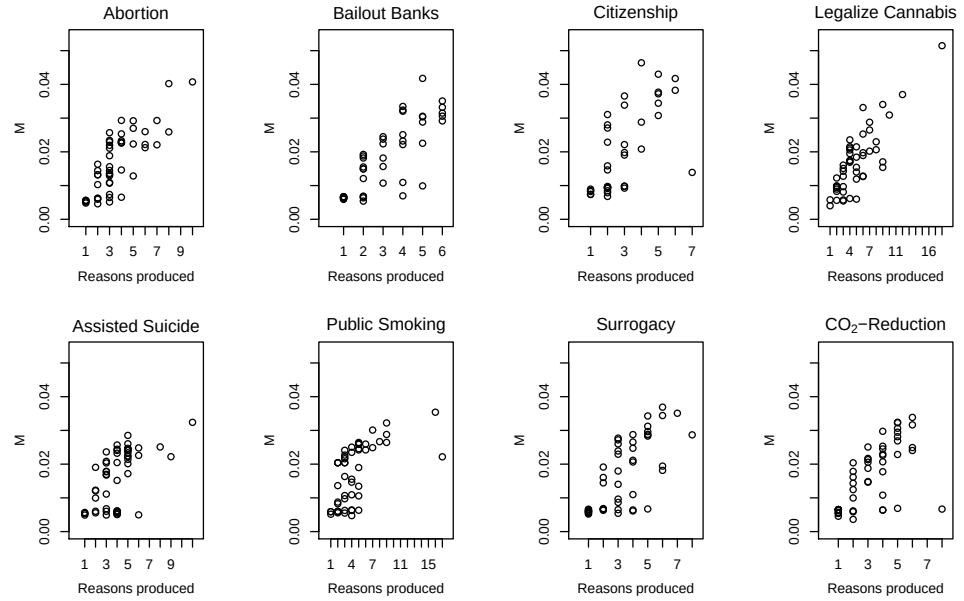

Figure 9. Mediation centrality M of an individual (y-axes) plotted as a function of the number of reasons produced by that individual, separately for each policy issue.

the most reasons produced, this not always the case. Bailing out banks, shortened naturalization, forced CO_2 reduction, public smoking ban, and surrogate legalization demonstrate cases where producing the most reasons does not make one the best mediator.

Figures 10 compares mediation centrality M and mediation representativeness, $\hat{\rho}$. The results are consistent with those shown in the simulations, indicating that individuals with higher $\hat{\rho}$ also have higher M. However, they also show how in this real-world context $\hat{\rho}$ can differ for individuals with the same M, such as the outliers in abortion and citizenship, which represent dyads separate from the respective giant components (networks not shown). Figure 11 shows similar results when comparing mediation centrality with the residence times of 1,000 random walkers starting at each individual in each subgraph. M is strongly correlated with

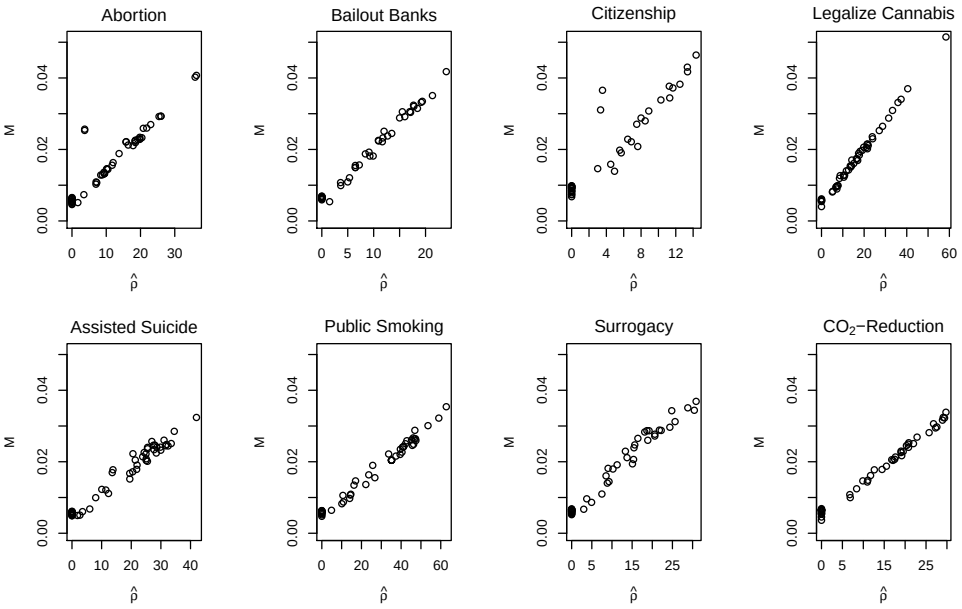

Figure 10. Mediation centrality M of an individual (y-axes) plotted as a function of mediation representativeness $\hat{\rho}$ of that individual, separately for each policy issue.

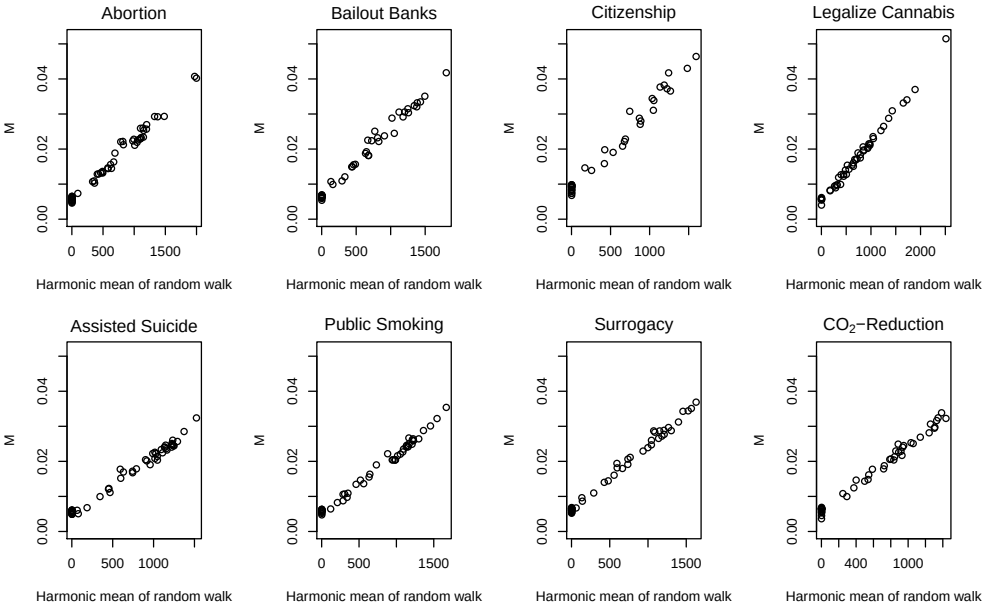

Figure 11. Mediation centrality M of an individual (y-axes) plotted as a function of the harmonic mean of random walk residence times (across both pro and con subgraphs), separately for each policy issue.

the residence times of random walkers in the reason space. These results for mediation representativeness and random walk residence times demonstrate that mediation centrality conforms to our intuition of what a good mediator for these policy issues might look like: Someone who is a best-recognizer-of-best-recognized reasons.

General Discussion

The present article has two goals. The first and primary goal is to introduce a new measure for network scientists that captures an interesting and useful property of bipartite adversarial policy networks. As we show, mediation centrality has useful quantitative properties that can identify nodes in bipartite networks that may be particularly suited for certain tasks

in adversarial settings. The second goal is to produce a measure of mediation that may be useful to social and cognitive scientists. Though this article focuses primarily on the former of these two goals, we nonetheless empirically demonstrated how mediation centrality is a meaningful measure for adversarial policy networks which should be useful in future studies that aim to provide more detailed quantification of the often rather qualitative conceptualization of mediation (see Deutsch et al., 2006).

In relation to complex networks, the mediation centrality measure we present, focusing on PageRank, is a particular instance of a more general family of possible measures of mediation centrality. Other centrality measures, such as closeness centrality or betweenness centrality, could be used in lieu of PageRank in the measure of mediation centrality we presented. However, we argue that for the task of mediation in adversarial policy networks, a measure of mediation that closely corresponds to social and cognitive processes is preferable to process-agnostic, descriptive measures, such as representativeness or other, more generic measures of network centrality. Nevertheless, such alternative measures are likely to be meaningful in other settings where they may closely correspond to other processes of interest. For example, when a bee colony tries to locate a position for their beehive that appropriately minimizes travel to resources of different types, closeness mediation centrality may be highly appropriate. Mediation centrality can also be adapted to weighted subgraphs in relation to their relative importance by some other criteria, such as, for example, the number of people holding a particular position on the issue, or the value of different reasons (e.g., pollen over nectar in the beehive example above; Seeley, 2009). Indeed, mediation centrality is a highly flexible approach for constructing quantitative measures and there is ample room for variations. For example, mediation centrality as proposed here is designed to measure mediation within a network dedicated to a given adversarial issue. But it may be valuable in the future to be able to quantify the degree of mediation across multiple policy issues, with correspondingly different associative structures.

In addition, future investigations of mediation from a more social psychological perspective should focus on several factors that were not addressed here. Foremost, we did not capture the strength with which individuals held the positions they reported themselves as having. For example, we did not capture the strength with which an individual supported laws against public smoking, only that they did or did not. It would be useful to have a more graded measure of position, as one then could investigate whether individuals with high mediation centrality are also individuals who hold more moderate positions on these issues. This, of course, may not be the case. Good mediators by our measure may also—by virtue of their knowledge of what other individuals believe—be better persuaders. In this respect, future studies of mediation would also benefit by investigating the extent to which individuals with high mediation centrality can produce arguments that are more likely to lead to solutions recognized by both sides.

Future studies should also investigate the extent to which individuals believe the reasons they acknowledge to reflect legitimate arguments. Although we use a rather coarse measure of acknowledgement, merely involving the production of a reason, it may be that these reasons are acknowledged to different degrees. Both of the above issues could be adapted into future measures of mediation. Finally, it is important to note that mediation centrality, as we propose it here, is social-network agnostic. It solely focuses on *what you know* and not on *who you know*. Nonetheless, in many contexts, mediators may be most effective when they simultaneously know the relevant parties involved *and* recognize the best-recognized reasons held on alternate sides of the issue.

Conclusion

Individuals who can take the perspectives held by opposing sides of an issue and frame arguments in a way that both sides can agree on often help to generate better solutions during conflict resolution (Drake & Donohue, 1996; Galinsky et al., 2008; Kemp & Smith, 1994). We apply this concept to mediators by extending perspective taking to a process-based account of mediation. This allows us to introduce mediation centrality, a metric for quantifying the mediation value of individuals in adversarial policy networks. Mediation centrality formalizes the notion of a good mediatior in a collective cognitive representation of an adversarial policy space aggregated across multiple individuals and positions. Using simulations and empirical data from eight real-world policy issues, we show that mediation centrality follows the intuition of what it means to be a good mediator, and we further show how this outperforms other measures and captures the logic of a random walk over reason space.

Data Availability

The data and code used to support the findings of this study have been deposited in an Open Science Framework project (https://osf.io/9a6wj).

Conflicts of Interest

The authors declare no conflicts of interest.

Funding Statement

This work was supported by the Royal Society Wolfson Research Merit Award (WM160074) and a Fellowship from the Alan Turing Institute (to TH).

Acknowledgments

We thank Carmen Kaiser for coding the experiment, Theresa Schmitt for documenting the experiment, the CDS research assisants for recruiting participants, and Dania Esch, Eva Günther, Sebastian Lucht, and Sarah Turowski for coding participants' responses.

References

Austerweil, J. L., Abbott, J. T., & Griffiths, T. L. (2012). Human memory search as a random walk in a semantic network. In *Advances in neural information processing systems* (pp. 3041–3049).

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531–1542.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71.

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: Structure and dynamics. *Entropy*, *12*(5), 1264–1302.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*(1-7), 107–117. doi: 10.1016/S0169-7552(98)00110-X

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113.

Deutsch, M., Coleman, P. T., & Marcus, E. C. (2006). *The handbook of conflict resolution: Theory and practice* (2nd ed.). San Francisco, CA: Jossey-Bass.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, *60*(11), 2229–2243.

Drake, L. E., & Donohue, W. A. (1996). Communicative framing theory in conflict resolution. *Communication Research*, *23*(3), 297–322.

Estrada, E., & Rodriguez-Velazquez, J. A. (2005). Subgraph centrality in complex networks. *Physical Review E*, *71*(5), 056103.

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*, 563–588.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

Galinsky, A. D., Maddux, W. W., Gilin, D., & White, J. B. (2008). Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological Science*, *19*(4), 378–384.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science*, *18*(12), 1069–1076.

Hills, T. T. (2018). The dark side of information proliferation. *Perspectives on Psychological Science*. (Advanced online publication.) doi: 10.1177/1745691618803647

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431.

Hills, T. T., & Segev, E. (2014). The news is American but our memories are... Chinese? *Journal of the Association for Information Science and Technology*, *65*(9), 1810–1819.

Hobbes, T. (1998). *Leviathan*. Oxford, England: Oxford University Press. (Original work published 1651).

Kemp, K. E., & Smith, W. P. (1994). Information exchange, toughness, and integrative bargaining: The roles of explicit cues and perspective-taking. *International Journal of Conflict Management*, *5*(1), 5–21.

Ku, G., Wang, C. S., & Galinsky, A. D. (2015). The promise and perversity of perspective-taking in organizations. *Research in Organizational Behavior*, *35*, 79–102.

Mutz, D. C. (2002). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, *96*(1), 111-126.

Newman, M. (2018). *Networks*. Oxford University Press.

Nikolov, D., Oliveira, D. F., Flammini, A., & Menczer, F. (2015). Measuring online social bubbles. *PeerJ Computer Science*, *1*, e38.

Norris, J. R. (1998). *Markov chains* (No. 2). Cambridge University Press.

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. Retrieved from `http://hdl.handle.net/20.500.11780/3343`

Russell, T., & Reimer, T. (2018). Using semantic networks to define the quality of arguments. *Communication Theory*, *28*(1), 46–68.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603.

Seeley, T. D. (2009). *The wisdom of the hive: The social physiology of honey bee colonies*. Harvard University Press.

Shi, F., Shi, Y., Dokshin, F. A., Evans, J. A., & Macy, M. W. (2017). Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nature Human Behaviour*, *1*(4), 0079.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Scientific Reports*, *2*, 335.

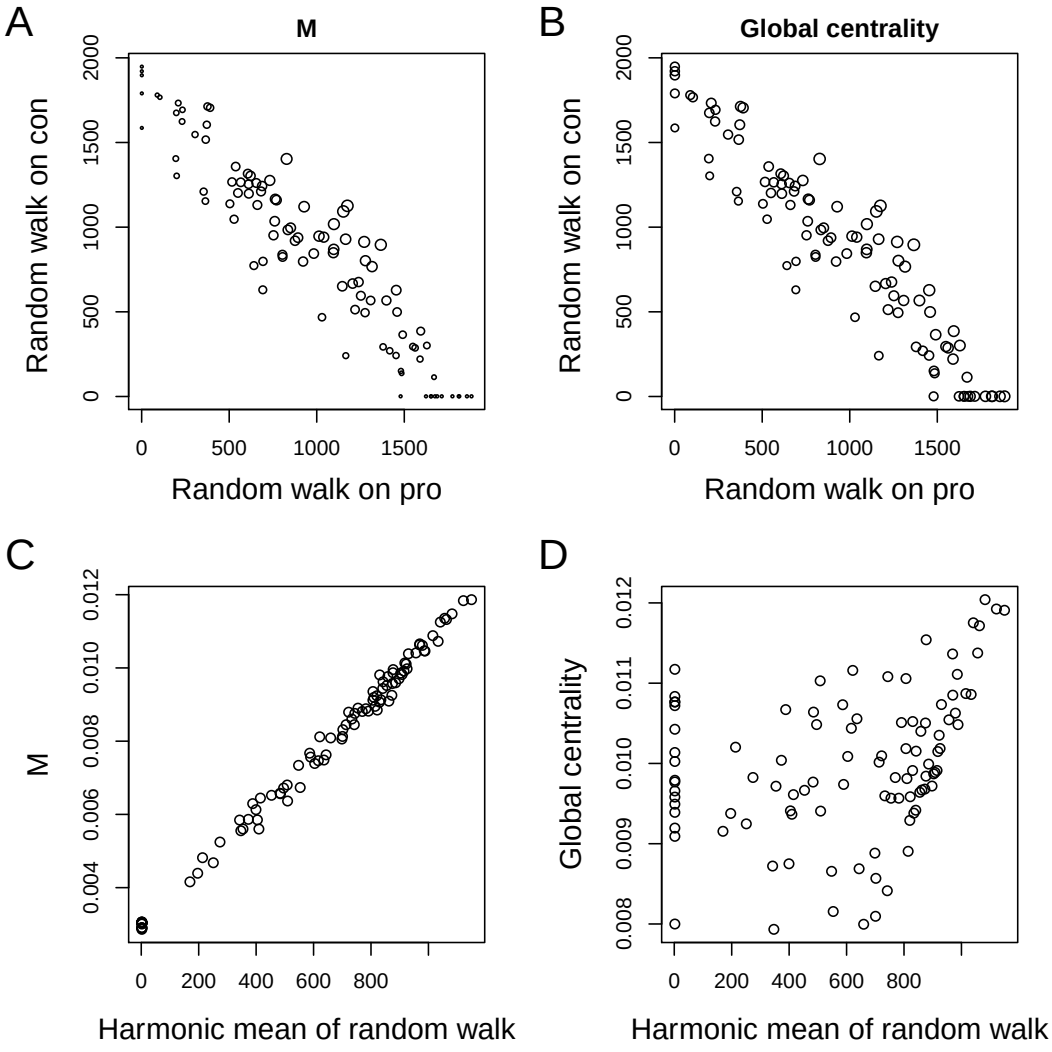

Figure A1. Comparison of mediation centrality against random walk residence times in Simulation 1. Otherwise this figure follows Figure 4.

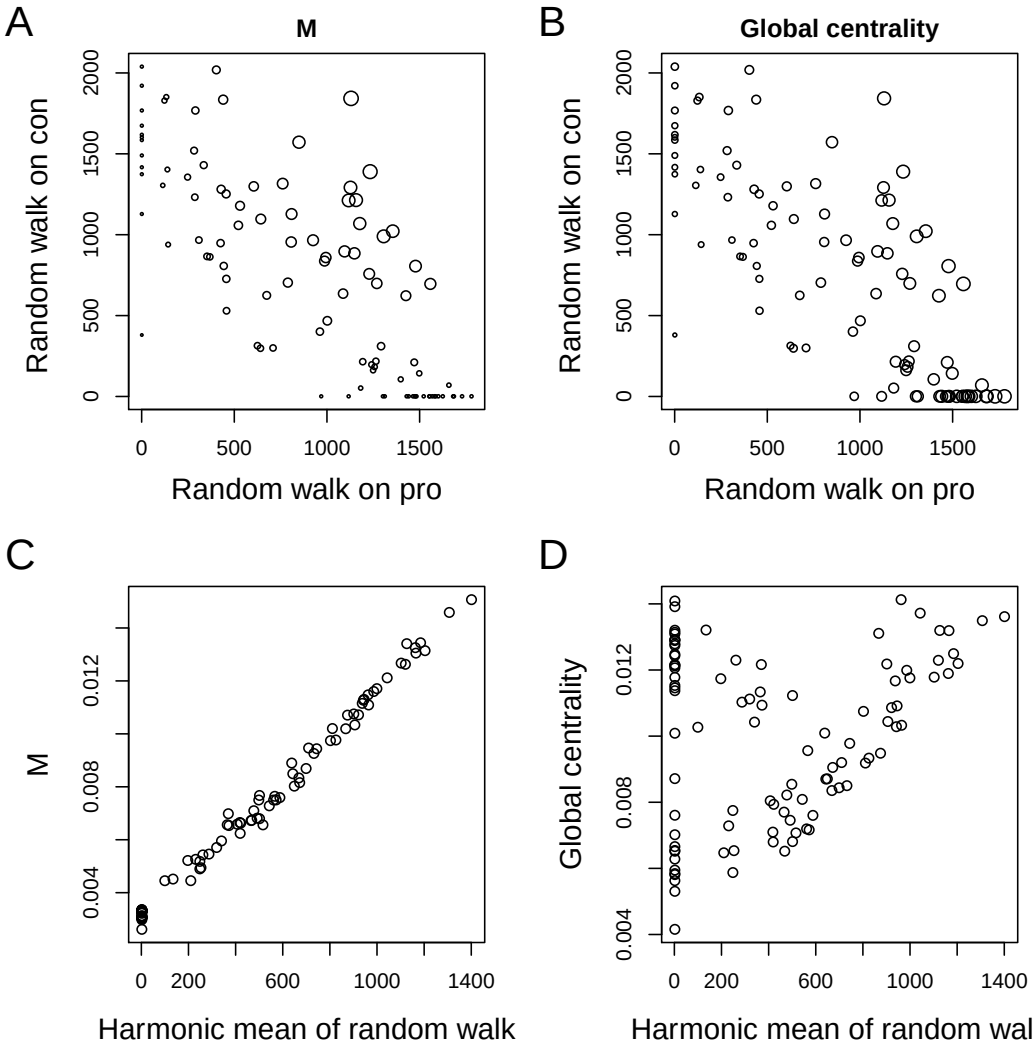

Figure A2. Comparison of mediation centrality against random walk residence times in Simulation 2. Otherwise this figure follows Figure 6.